LEAST-SQUARES FINITE ELEMENT METHODS

Pavel Bochev[1,2] and Max Gunzburger[3]

The root cause for the remarkable success of early finite element methods (FEMs) is their intrinsic connection with Rayleigh-Ritz principles. Yet, many partial differential equations (PDEs) are not associated with unconstrained minimization principles and give rise to less favorable settings for FEMs. Accordingly, there have been many efforts to develop FEMs for such PDEs that share some, if not all, of the attractive mathematical and algorithmic properties of the Rayleigh-Ritz setting. Least-squares principles achieve this by abandoning the naturally occurring variational principle in favor of an artificial, external energy-type principle. Residual minimization in suitable Hilbert spaces defines this principle. The resulting least-squares finite element methods (LSFEMs) consistently recover almost all of the advantages of the Rayleigh-Ritz setting over a wide range of problems and, with some additional effort, they can often create a completely analogous variational environment for FEMs.

A more detailed presentation of least squares finite element methods is given in [1].

Abstract LSFEM theory. Consider the abstract PDE problem

$$\text{(1)} \qquad \textit{find } u \in X \quad \textit{such that} \quad \mathcal{L}u = f \quad \textit{in } Y,$$

where X and Y are Hilbert spaces, $\mathcal{L} : X \mapsto Y$ is a bounded linear operator, and $f \in Y$ is given data. Assume (1) to be well posed so that there exist positive constants α and β such that

$$\text{(2)} \qquad \beta\|u\|_X \leq \|\mathcal{L}u\|_Y \leq \alpha\|u\|_X \qquad \forall\, u \in X.$$

The *energy balance* (2) is the starting point in the development of LSFEMs. It gives rise to the unconstrained minimization problem, i.e., the *least-squares principle* (LSP),

$$\text{(3)} \qquad \{J, X\} \;\rightarrow\; \left\{ \min_{u\in X} J(u; f), \quad J(u; f) = \|\mathcal{L}u - f\|_Y^2 \right\},$$

where $J(u, f)$ is the *residual energy* functional. From (2), it follows that $J(\cdot;\cdot)$ is *norm equivalent*:

$$\text{(4)} \qquad \beta^2\|u\|_X^2 \leq J(u; 0) \leq \alpha^2\|u\|_X^2 \qquad \forall\, u \in X.$$

Norm equivalence (4) and the Lax-Milgram Lemma imply that the Euler-Lagrange equation of (3):

$$\text{(5)} \qquad \textit{find } u \in X \quad \textit{such that} \quad \big(\mathcal{L}v, \mathcal{L}u\big)_Y \equiv Q(u, w) = F(w) \equiv \big(\mathcal{L}v, f\big)_Y \quad \forall\, w \in X$$

is well-posed because $Q(u, w)$ is an equivalent inner product on $X \times X$. The unique solution of (5), resp. (3), coincides with the solution of (1).

We define an LSFEM by restricting (3) to a family of finite element subspaces $X^h \subset X$, $h \to 0$. The LSFEM approximation $u^h \in X^h$ to the solution $u \in X$ of (1) or (3) is the solution of the unconstrained minimization problem

$$\text{(6)} \qquad \{J, X^h\} \;\rightarrow\; \left\{ \min_{u^h\in X^h} J(u^h; f), \quad J(u; f) = \|\mathcal{L}u^h - f\|_Y^2 \right\}.$$

To compute u^h, we solve the Euler-Lagrange equation corresponding to (6):

$$\text{(7)} \qquad \textit{find } u^h \in U^h \quad \textit{such that} \quad Q(u^h, w^h) = F(w^h) \qquad \forall\, w^h \in W^h.$$

---

[1]Numerical Analysis and Applications, Sandia National Laboratories, MS1320, Albuquerque, New Mexico, 87185-1320, USA.

[2]Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energys National Nuclear Security Administration under contract DE-AC04-94AL85000

[3]Department of Scientific Computing, 400 Dirac Science Library, Florida State University, Tallahassee Florida 32306-4120, USA.

Let $\{\phi_j^h\}_{j=1}^N$ denote a basis for X^h so that $u^h = \sum_{j=1}^N \vec{u}_j^h \phi_j^h$. Then, the problem (7) is equivalent to the linear system of algebraic equations

$$\mathbb{Q}^h \vec{u}^h = \vec{f}^h \tag{8}$$

for the unknown vector $\vec{u}^h$, where $\mathbb{Q}_{ij}^h = \big(\mathcal{L}\phi_j^h, \mathcal{L}\phi_i^h\big)_Y$ and $\vec{f}_i^h = \big(\mathcal{L}\phi_i, f\big)_Y$.

Theorem 1. *Assume that* (2)*, or equivalently,* (4)*, holds and that $X^h \subset X$. Then,*

– *the bilinear form $Q(\cdot,\cdot)$ is continuous, symmetric, and strongly coercive*
– *the linear functional $F(\cdot)$ is continuous*
– *the problem* (5) *has a unique solution $u \in X$ that is also the unique solution of* (3)
– *the problem* (7) *has a unique solution $u^h \in X^h$ that is also the unique solution of* (6)
– *the LSFEM approximation u^h is optimally accurate with respect to solution norm $\|\cdot\|_X$ for which* (1) *is well posed, i.e., for some constant $C > 0$*

$$\|u - u^h\|_X \le C \inf_{v^h \in X^h} \|u - v^h\|_X \tag{9}$$

– *the matrix $\mathbb{Q}^h$ of* (8) *is symmetric and positive definite.* □

Theorem 1 only assumes that (1) *is well posed and that X^h is conforming.* It does not require $\mathcal{L}$ to be positive self-adjoint as it would have to be in the Rayleigh-Ritz setting, nor does it impose any compatibility conditions on X^h that are typical of other FEMs. Despite the generality allowed for in (1), *the LSFEM based on* (6) *recovers* all *the desirable features possessed by finite element methods in the Rayleigh-Ritz setting.* This is what makes LSFEMs intriguing and attractive.

Practical LSFEM. Intuitively, a "practical" LSFEM has coding complexity and conditioning comparable to that of other FEMs for the same PDE. The LSP $\{J, X\}$ in (3) recreates a true Rayleigh-Ritz setting for (1), yet the LSFEM $\{J, X^h\}$ in (6) may be impractical. Thus, sometimes it is necessary to replace $\{J, X\}$ by a practical discrete alternative $\{J^h, X^h\}$. Two opposing forces affect the construction of $\{J^h, X^h\}$: a desire to keep the resulting LSFEM simple, efficient, and practical and a desire to recreate the true Rayleigh-Ritz setting. The latter requires J^h to be as close as possible to the "ideal" norm-equivalent setting in (3).

The transformation of $J(\cdot,\cdot)$ into a discrete functional $J^h(\cdot,\cdot)$ illustrates the interplay between these issues. To this end, it is illuminating to write the energy balance (2) in the form

$$C_1 \|\mathcal{S}_X u\|_0 \le \|\mathcal{S}_Y \circ \mathcal{L} u\|_0 \le C_2 \|\mathcal{S}_X u\|_0\,, \tag{10}$$

where $\mathcal{S}_X, \mathcal{S}_Y$ are norm-generating operators for X, Y, respectively, with $L^2(\Omega)$ acting as a pivot space. At the least, practicality requires that the basis of X^h can be constructed with no more difficulty than for Galerkin FEM for the same PDE. To secure this property we ask that the domain $D(\mathcal{S}_X)$ of $\mathcal{S}_X$ contains "practical" discrete subspaces. Transformation of (1) into an *equivalent first-order system* PDE achieves this. Then, practicality of the "ideal" LSFEM (6) depends solely on the effort required to compute $\mathcal{S}_Y \circ \mathcal{L} u^h$. If this effort is deemed reasonable, the original energy norm $|||u||| = \|\mathcal{S}_Y \circ \mathcal{L} u\|_0$ can be retained and the transition process is complete. Otherwise, we proceed to replace the composite operator $\mathcal{S}_Y \circ \mathcal{L}$ by a computable discrete approximation $\mathcal{S}_Y^h \circ \mathcal{L}^h$. We may need a *projection* operator π^h that maps the data f to the domain of $\mathcal{S}_Y^h$. The conversion process and the key properties of the resulting LSFEM can be encoded by the *transition* diagram

$$\begin{array}{ccccc} J(u;f) & = & \|\mathcal{S}_Y \circ (\mathcal{L}u - f)\|_0^2 & \to & |||u||| \\ \downarrow & & \downarrow & & \downarrow \\ J^h(u^h;f) & = & \|\mathcal{S}_Y^h \circ (\mathcal{L}^h u^h - \pi^h f)\|_0 & \to & |||u^h|||_h \end{array} \tag{11}$$

and the companion *norm-equivalence* diagram

$$\begin{array}{ccccc} C_1 \|u\|_X & \le & |||u||| & \le & C_2 \|u\|_X \\ \downarrow & & \downarrow & & \downarrow \\ C_1(h) \|u^h\|_X & \le & |||u^h|||_h & \le & C_2(h) \|u^h\|_X . \end{array} \tag{12}$$

Because $\mathcal{L}$ defines the problem being solved, the choice of $\mathcal{L}^h$ governs the accuracy of the LSFEM. The goal here is to make J^h as close as possible to J for the exact solution of (1). On the other hand, $\mathcal{S}_Y$ defines the energy balance of (1), i.e., the proper scaling between data and solution. As a result, the main objective in the choice of $\mathcal{S}_Y^h$ is to ensure that the scaling induced by J^h is as close as possible to (2), i.e., to "bind" the LSFEM to the energy balance of the PDE.

Taxonomy of LSFEMs. Assuming that X^h is practical, restriction of $\{J, X\}$ to X^h transforms (3) into the *compliant* LSFEM $\{J, X^h\}$ in (6). Apart from this "ideal" LSFEM which reproduces the classical Rayleigh-Ritz principle, there are two other kinds of LSFEMs that gradually drift away from this setting, primarily by *simplifying the approximations* of the norm-generating operator $\mathcal{S}_Y$. Mesh-independent $C_1(h)$ and $C_2(h)$ in (12) characterize the *norm-equivalent* class, which retains virtually all attractive properties of the Rayleigh-Ritz setting, including identical convergence rates and matrix condition numbers. A mesh-dependent norm-equivalence (12) distinguishes the *quasi-norm-equivalent* class, which admits the broadest range of LSFEMs, but can give problems with higher condition numbers.

Examples. We use the Poisson equation for which $\mathcal{L} = -\Delta$ to illustrate different classes of LS-FEMs. One energy balance (2) for this equation corresponds to $X = H^2(\Omega) \cap H_0^1(\Omega)$ and $Y = L^2(\Omega)$:

$$\alpha\|u\|_2 \le \|\Delta u\|_0 \le \beta\|u\|_2.$$

The associated LSP

$$\{J, X\} \;\to\; \left\{ \min_{u \in X} J(u; f),\;\; J(u; f) = \|\Delta u - f\|_0^2 \right\}$$

leads to impractical LSFEMs because finite element subspaces of $H^2(\Omega)$ are not easy to construct.

Transformation of $-\Delta u = f$ into the equivalent first-order system

$$\nabla \cdot \mathbf{q} = f \quad \text{and} \quad \nabla u + \mathbf{q} = 0 \tag{13}$$

can solve this problem. The spaces $X = H_0^1(\Omega) \times [L^2(\Omega)]^d$, $Y = H^{-1}(\Omega) \times [L^2(\Omega)]^d$ have practical finite element subspaces and provide the energy balance

$$\alpha(\|u\|_1 + \|\mathbf{q}\|_0) \le \|\nabla \cdot \mathbf{q}\|_{-1} + \|\nabla u + \mathbf{q}\|_0 \le \beta(\|u\|_1 + \|\mathbf{q}\|_0).$$

This energy balance gives rise to the *minus one norm* LSP

$$\{J, X\} \;\to\; \left\{ \min_{(u,\mathbf{q}) \in X} J(u, \mathbf{q}; f), \quad J(u, \mathbf{q}; f) = \|\nabla \cdot \mathbf{q} - f\|_{-1}^2 + \|\nabla u + \mathbf{q}\|_0^2 \right\}. \tag{14}$$

However, (14) is still impractical because the norm-generating operator $\mathcal{S}_{H^{-1}} = (-\Delta)^{-1/2}$ is not computable in general. The simple approximation $\mathcal{S}_{H^{-1}}^h = h\mathbf{I}$ yields the *weighted* LSFEM

$$\{J^h, X^h\} \;\to\; \left\{ \min_{(u^h,\mathbf{q}^h) \in X^h} J^h(u^h, \mathbf{q}^h; f), \quad J^h(u^h, \mathbf{q}^h; f) = h^2\|\nabla \cdot \mathbf{q}^h - f\|_0^2 + \|\nabla u^h + \mathbf{q}^\mathbf{h}\|_0^2 \right\} \tag{15}$$

which is quasi-norm equivalent. The more accurate approximation $\mathcal{S}_{H^{-1}}^h = h\mathbf{I} + \mathbf{K}^{h^{1/2}}$, where $\mathbf{K}^h$ is a spectrally-equivalent preconditioner for $-\Delta$ gives the *discrete minus-one norm* LSFEM

$$\{J^h, X^h\} \;\to\; \left\{ \min_{(u^h,\mathbf{q}^h) \in X^h} J^h(u^h, \mathbf{q}^h; f), \quad J^h(u^h, \mathbf{q}^h; f) = \|\nabla \cdot \mathbf{q}^h - f\|_{-h}^2 + \|\nabla u^h + \mathbf{q}^h\|_0^2 \right\} \tag{16}$$

which is norm-equivalent.

The first-order system (13) also has the energy balance

$$\alpha(\|u\|_1 + \|\mathbf{q}\|_{div}) \le \|\nabla \cdot \mathbf{q}\|_0 + \|\nabla u + \mathbf{q}\|_0 \le \beta(\|u\|_1 + \|\mathbf{q}\|_{div})$$

which corresponds to $X = H_0^1(\Omega) \times H(div, \Omega)$ and $Y = L^2(\Omega) \times [L^2(\Omega)]^d$. The associated LSP

$$\{J, X\} \;\to\; \left\{ \min_{(u,\mathbf{q}) \in X} J(u, \mathbf{q}; f), \quad J(u, \mathbf{q}; f) = \|\nabla \cdot \mathbf{q} - f\|_0^2 + \|\nabla u + \mathbf{q}\|_0^2 \right\} \tag{17}$$

is practical. Approximation of the scalar u by standard nodal elements and of the vector $\mathbf{q}$ by div-conforming elements, such as Raviart-Thomas, BDM, or BDFM, yields a compliant LSFEM which under some conditions has the exact same local conservation property as the mixed Galerkin method for (13).

References

[1] P. Bochev and M. Gunzburger, *Least Squares Finite Element Methods*, Springer, Berlin, 2009.